Correlation between Host-Guest Binding and Host Amplification in Simulated Dynamic Combinatorial Libraries

Corbett, Peter T.; Otto, Sijbren; Sanders, Jeremy K.M.

Published in:
Chemistry

DOI:
10.1002/chem.200400300

Publication date:
2004

Chem. Eur. J. **2004**

Supporting Information

for

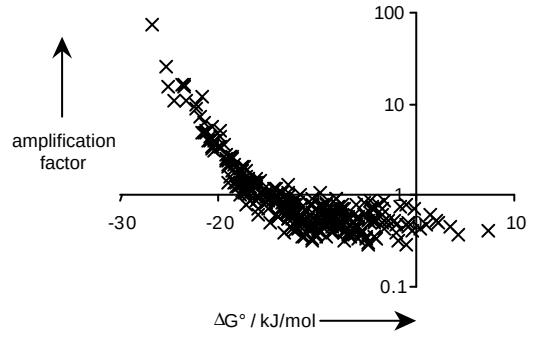
Correlation between host-guest binding and host amplification in simulated dynamic combinatorial libraries

By

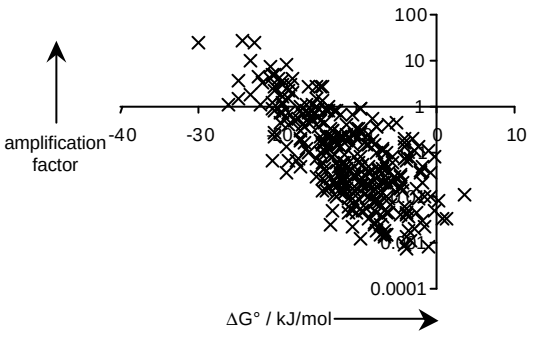
Peter T. Corbett, Sijbren Otto* and Jeremy K. M. Sanders*

1) Examples across the observed range of R^2 values for the correlation between amplification factor and binding strength

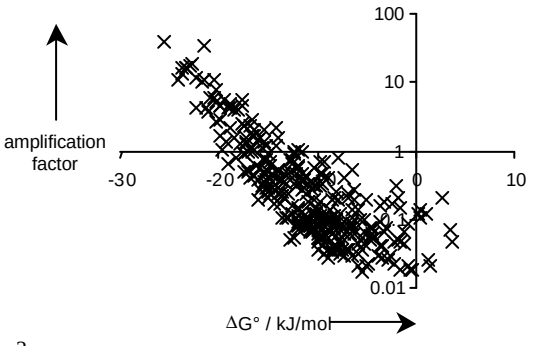
The following graphs show the correlation between amplification factor and the free energy of binding in four simulated DCLs – all were taken from the dataset described in the main text, and selected to represent a range of R^2 values (calculated as described in the main text). Each graph represents a different DCL, with a different set of randomly-generated binding constants.



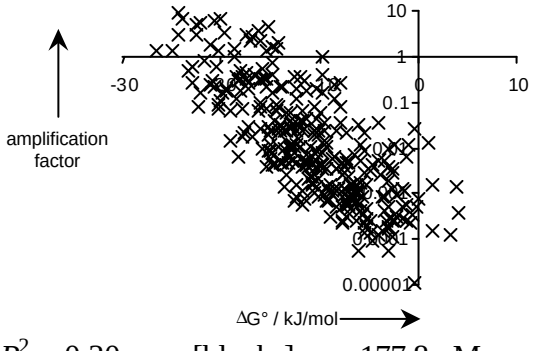
R^2 = 0.90 [blocks]$_{total}$ = 10mM
[template] = 3.2mM



R^2 = 0.50 [blocks]$_{total}$ = 56.2mM
[template] = 56.2mM



R^2 = 0.70 [blocks]$_{total}$ = 31.6mM
[template] = 31.6mM



R^2 = 0.30 [blocks]$_{total}$ =177.8mM
[template] = 316.2mM

2) Method for calculations

The simulation of dynamic combinatorial libraries was performed by the computer program DCLSim, which randomly generates a set of equilibrium constants and assigns them to a set of receptors, and then calculates the distribution of products at equilibrium based on these equilibrium constants[1] and a user-specified set of initial concentrations of the building blocks and templates.

DCLSim simulates the many equilibria in DCLs by considering the reversible reactions between building blocks blocks.[2] Consider a set of building blocks, A, B, C etc., and a set of compounds, $A_xB_yC_z\ldots$, $A_uB_vC_w\ldots$, etc., giving rise to a set of chemical equations as below:

$$xA + yB + zC + \ldots \rightleftharpoons A_xB_yC_z\ldots \quad (K_{xyz})$$

The equilibrium constants for the formation of library members from the building blocks are generated by simply counting the number of different sequences of building blocks that correspond to the same composition. For example, the library member ABCD can be represented by 4! = 24 sequences, so it would be given an equilibrium constant of 24. This method gives the desired preference for heterooligomers, and for small oligomers at sub-molar concentrations.

Binding to the template is simulated by including the template as another building block, and the set of host-guest complexes between the library members and the template as further library members. The equilibrium constants for the formation of these are generated by multiplying the equilibrium constants for the host compounds by the randomly generated binding constant. After the equilibrium is calculated, the concentrations of the host-guest complexes and the free hosts are added together, to give the final host concentrations.

A set of equations of the form of Equation 1 can be written to describe the mass balance in the system, and a set of equations in the form of Equation 2 (a re-arrangement of the equilibrium constant expression) can be written to describe the formation of the compounds from the building blocks.

$$[A]_{init} = [A]_{free} + x[A_xB_yC_z\ldots] + u[A_uB_vC_w\ldots] + \ldots \qquad \text{(Equation 1)}$$

$$[A_xB_yC_z\ldots] = K_{xyz}[A]_{free}^{x}[B]_{free}^{y}[C]_{free}^{z}\ldots \qquad \text{(Equation 2)}$$

By solving these two sets of simultaneous equations, all of the relevant concentrations can be found. This is done using an iterative method. Estimates of the free concentrations of the building blocks are made – in the first round, these estimates are simply the specified total building block concentrations. From these, the concentrations of the other compounds are calculated, by substituting them into Equation 2. These concentrations are then used to calculate the total building block concentrations, which are compared to the specified total concentrations. If these are all within 0.001%, then all of the concentrations are accepted. Otherwise, new estimates of the free concentration are made using Equation 3, taken from the literature[2]:

$$[A]_{free,improved} = \frac{[A]_{free,old}}{\sqrt{[A]_{init,calculated} / [A]_{init,specified}}} \qquad \text{(Equation 3)}$$

During the development of DCLSim, it was found that simply using this procedure sometimes led to an oscillatory cycle instead of convergence. This problem was overcome by modifying the value of $[A]_{free,\ improved}$ using Equation 4. The parameter *c* is 1 on the first iteration, 2 on the second, and increases by 1 on any round where it is detected that a newly calculated total concentration is further from the specified total than the old one.

$$[A]_{free,damped} = \frac{([A]_{free,old})(c-1) + [A]_{free,improved}}{c} \qquad \text{(Equation 4)}$$

DCLSim also allows for the simulation of equilibria that do not include the building blocks that make up the other compounds as free species (i.e. free building blocks). For example, in a DCL of macrocyclic disulfides, no free thiols (or only immeasurably small traces) are left in the final equilibrated solution. Although this situation may not seem amenable to the approach of calculating everything based on the concentration of free building blocks, nevertheless these equilibria can be simulated by using a modification of the algorithm, such that the *free* concentrations of the building blocks are not included in their calculated *total* concentrations. As such, the building blocks in the program do not represent real chemical species, but are convenient mathematical fictions, acting as 'virtual intermediates' in the complex equilibria between different complex species.

The validity of this method can be shown, thus: consider three sets of chemical species: D, E, F, etc.; I, J, K, etc. and N, O, P, etc. If it is possible to write two balanced equations for the following equilibria:

$$dD + eE + fF + \ldots \rightleftharpoons_{K_{1,2}} iI + jJ + kK + \ldots$$

$$dD + eE + fF + \ldots \rightleftharpoons_{K_{1,3}} nN + oO + pP + \ldots$$

then any set of concentrations that satisfies Equations 5 and 6 will also satisfy Equation 7.

$$K_{1,2} = \frac{[I]^i [J]^j [K]^k \ldots}{[D]^d [E]^e [F]^f \ldots} \qquad \text{(Equation 5)}$$

$$K_{1,3} = \frac{[N]^n [O]^o [P]^p \ldots}{[D]^d [E]^e [F]^f \ldots} \qquad \text{(Equation 6)}$$

$$\frac{K_{1,3}}{K_{1,2}} = \frac{[N]^n [O]^o [P]^p \ldots}{[I]^i [J]^j [K]^k \ldots} \qquad \text{(Equation 7)}$$

$K_{1,3}/K_{1,2}$ fits the form of the equilibrium constant for the following process:

$$iI + jJ + kK + \ldots \rightleftharpoons_{K_{2,3}} nN + oO + pP + \ldots$$

$$K_{2,3} = \frac{[N]^n [O]^o [P]^p \ldots}{[I]^i [J]^j [K]^k \ldots} \qquad \text{(Equation 8)}$$

As a result of this, in any set of concentrations of D, E, F, I, J, K, N, O, P etc. that satisfy Equations 5 and 6, where $K_{1,3}/K_{1,2}$ is equal to $K_{2,3}$, the concentrations of I, J, K, N, O, P etc. will satisfy Equation 8. This applies even if compounds D, E, F etc. and the equilibria between them and the other compounds *are completely fictitious and arbitrarily chosen*. Thus, a set of concentrations generated to fit the fictitious equilibria will satisfy any real equilibria in the system of compounds described. Furthermore, it does not matter how the concentrations are generated, so long as they are checked to make sure that they fit the equilibria.

DCLSim was written by PTC – please contact any of the authors of this paper for availability.

[1] DCLSim may also be used to calculate distributions of products based upon user-specified sets of equilibrium constants.
[2] This method is a modification of the COGS subroutine of the COMICS program: D. D. Perrin, I. G. Sayce, *Talanta* **1967**, *14*, 833-842. [Some differences in the precise implementation of the algorithm were made, an insignificant distinction between metal ions and ligands was removed, as was the influence of pH, the concept of virtual intermediates was added and the method of improving the estimated free concentrations was improved.]